

2 - Modelo de Regresión Logística Multinomial

1. Contexto y Objetivo del Modelo

- **Estudio:** Factores asociados al estado nutricional de niños en edad preescolar (2-5 años) en comunidades rurales.
- **Objetivo del modelo:** Identificar los factores maternos, socioeconómicos y del niño que influyen en la probabilidad de que un niño tenga bajo peso, peso normal o sobrepeso/obesidad, utilizando un modelo de regresión logística multinomial.
- **Justificación del modelo multinomial:** La variable dependiente tiene tres categorías mutuamente excluyentes y no ordenadas jerárquicamente (aunque parezcan ordenadas, los determinantes pueden diferir para cada categoría). Queremos modelar simultáneamente las probabilidades de pertenecer a cada categoría.

Simulación de la data

```
## 'data.frame': 500 obs. of 12 variables:
## $ id : int 1 2 3 4 5 6 7 8 9 10 ...
## $ estado_nutricional: Factor w/ 3 levels "bajo_peso","normal",...: 2 3 1 2 2 2 2 1 2 ...
## $ edad_meses : num 34 52 39 56 58 26 43 56 44 40 ...
## $ sexo : Factor w/ 2 levels "femenino","masculino": 2 2 2 2 2 2 1 2 1 2 ...
## $ peso_nacimiento : num 2.7 3.5 2.4 4.1 4.1 3.2 3.9 2.7 2.2 3.1 ...
## $ edad_madre : num 43 45 30 24 37 33 34 30 30 34 ...
## $ educacion_madre : Factor w/ 3 levels "primaria","secundaria",...: 2 2 2 1 1 1 2 2 2 2 ...
## $ imc_madre : num 32.5 29.3 30.5 23.4 29.4 25.9 29.1 31.4 33.3 27 ...
## $ ingreso_familiar : num 406 1143 579 826 384 ...
## $ zona : Factor w/ 2 levels "rural","urbana": 2 2 1 2 2 2 2 2 1 1 ...
## $ lactancia_materna : num 7 20 14 19 7 3 21 0 14 23 ...
## $ dietas_solidos : Factor w/ 2 levels "no","si": 1 1 1 1 1 1 1 1 2 2 ...

## id estado_nutricional edad_meses sexo
## Min. : 1.0 bajo_peso:150 Min. :24.00 femenino :248
## 1st Qu.:125.8 normal :290 1st Qu.:33.00 masculino:252
## Median :250.5 sobrepeso: 60 Median :41.00
## Mean :250.5 Mean :41.83
## 3rd Qu.:375.2 3rd Qu.:50.00
## Max. :500.0 Max. :60.00
## peso_nacimiento edad_madre educacion_madre imc_madre
## Min. :2.000 Min. :18.00 primaria :192 Min. :18.10
## 1st Qu.:2.600 1st Qu.:25.00 secundaria:232 1st Qu.:22.27
## Median :3.200 Median :31.00 superior : 76 Median :26.60
## Mean :3.235 Mean :31.57 Mean :26.61
## 3rd Qu.:3.800 3rd Qu.:38.00 3rd Qu.:30.90
## Max. :4.500 Max. :45.00 Max. :35.00
## ingreso_familiar zona lactancia_materna dietas_solidos
## Min. : 201.0 rural :206 Min. : 0.00 no:358
## 1st Qu.: 450.8 urbana:294 1st Qu.: 6.00 si:142
## Median : 704.0 Median :12.00
## Mean : 700.7 Mean :11.77
## 3rd Qu.: 953.8 3rd Qu.:17.00
## Max. :1199.0 Max. :24.00
```

##	id	estado_nutricional	edad_meses	sexo	peso_nacimiento	edad_madre
## 1	1	normal	34	masculino	2.7	43
## 2	2	sobrepeso	52	masculino	3.5	45
## 3	3	bajo_peso	39	masculino	2.4	30
## 4	4	normal	56	masculino	4.1	24
## 5	5	normal	58	masculino	4.1	37
## 6	6	normal	26	masculino	3.2	33

##	educacion_madre	imc_madre	ingreso_familiar	zona	lactancia_materna
## 1	secundaria	32.5	406	urbana	7
## 2	secundaria	29.3	1143	urbana	20
## 3	secundaria	30.5	579	rural	14
## 4	primaria	23.4	826	urbana	19
## 5	primaria	29.4	384	urbana	7
## 6	primaria	25.9	859	urbana	3

##	dietas_solidos
## 1	no
## 2	no
## 3	no
## 4	no
## 5	no
## 6	no

Descripción de variables

Variable Dependiente:

- **estado_nutricional:** Categoría nutricional del niño (3 niveles)
 - bajo_peso: Peso/edad < -2 desviaciones estándar (Z-score OMS)
 - normal: Peso/edad entre -2 y +1 desviaciones estándar
 - sobrepeso: Peso/edad > +1 desviaciones estándar

Variables Independientes:

1. Características del niño:

- **edad_meses:** Edad en meses (continuo, 24-60 meses)
- **sexo:** Sexo biológico (categórico: **femenino**, **masculino**)
- **peso_nacimiento:** Peso al nacer en kg (continuo, 2.0-4.5 *kg*)

2. Características maternas:

- **edad_madre:** Edad de la madre en años (continuo, 18-45 años)
- **educacion_madre:** Nivel educativo (categórico: **primaria**, **secundaria**, **superior**)
- **imc_madre:** Índice de Masa Corporal materno (continuo, 18-35 *kg/m²*)

3. Factores socioeconómicos:

- **ingreso_familiar:** Ingreso mensual familiar en USD (continuo, 200-1200)
- **zona:** Zona de residencia (categórico: **urbana**, **rural**)

4. Prácticas de alimentación:

- **lactancia_materna:** Meses de lactancia materna exclusiva (continuo, 0-24 meses)
- **dietas_solidos:** Introducción de sólidos antes de 6 meses (categórico: **si**, **no**)

2. Análisis Exploratorio

Characteristic	Overall N = 500	bajo_peso N = 150	normal N = 290	sobrepeso N = 60	p-value
edad_meses	42 (10)	43 (10)	41 (10)	41 (11)	0.3
sexo					0.7
femenino	248 (50%)	78 (52%)	142 (49%)	28 (47%)	
masculino	252 (50%)	72 (48%)	148 (51%)	32 (53%)	
peso_nacimiento	3.24 (0.71)	3.17 (0.74)	3.25 (0.69)	3.33 (0.71)	0.3
edad_madre	32 (8)	33 (7)	31 (8)	31 (8)	0.021
educacion_madre					0.072
primaria	192 (38%)	61 (41%)	106 (37%)	25 (42%)	
secundaria	232 (46%)	60 (40%)	140 (48%)	32 (53%)	
superior	76 (15%)	29 (19%)	44 (15%)	3 (5.0%)	
imc_madre	26.6 (4.9)	26.7 (4.6)	26.5 (5.0)	27.1 (5.1)	0.7
ingreso_familiar	701 (287)	637 (290)	730 (281)	719 (290)	0.006
zona					0.008
rural	206 (41%)	77 (51%)	104 (36%)	25 (42%)	
urbana	294 (59%)	73 (49%)	186 (64%)	35 (58%)	
lactancia_materna	12 (7)	12 (6)	12 (7)	11 (7)	0.8
dietas_solidos					0.6
no	358 (72%)	103 (69%)	211 (73%)	44 (73%)	
si	142 (28%)	47 (31%)	79 (27%)	16 (27%)	

¹ Mean (SD); n (%)

² Kruskal-Wallis rank sum test; Pearson's Chi-squared test

Características generales de la muestra

- La muestra consta de 500 niños en edad preescolar (24-60 meses)
- **Distribución del estado nutricional:**
 - **58% (n=290)** tiene peso normal (categoría de referencia)
 - **30% (n=150)** presenta bajo peso
 - **12% (n=60)** tiene sobrepeso
- **Distribución equilibrada por sexo:** 50% femenino, 50% masculino

Diferencias significativas entre grupos

1. Variables con diferencias estadísticamente significativas ($p < 0.05$):

- **Edad de la madre ($p = 0.021$)**
 - Las madres de niños con **bajo peso** son significativamente mayores (33 ± 7 años)
 - Las madres de niños con **peso normal y sobrepeso** son más jóvenes (31 ± 8 años)
 - **Interpretación:** La mayor edad materna podría asociarse con prácticas alimentarias menos adecuadas o factores biológicos que predisponen al bajo peso infantil.

- **Ingreso familiar ($p = 0.006$)**
 - Familias de niños con **bajo peso** tienen menor ingreso (637 ± 290 USD)
 - Familias de niños con **peso normal** tienen mayor ingreso (730 ± 281 USD)
 - **Interpretación:** El nivel socioeconómico se asocia significativamente con el estado nutricional, siendo el bajo ingreso un factor de riesgo para el bajo peso.
- **Zona de residencia ($p = 0.008$)**
 - **51%** de niños con bajo peso viven en zona rural.
 - Solo **36%** de niños con peso normal viven en zona rural.
 - **Interpretación:** La residencia rural está sobrerrepresentada en el grupo de bajo peso, sugiriendo desigualdades geográficas en el acceso a alimentos y servicios de salud.

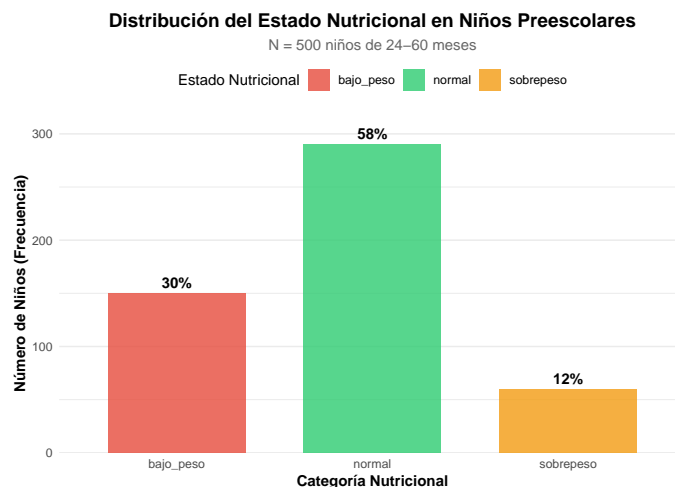
2. Variables con Diferencias Marginalmente Significativas ($p < 0.10$):

- **Educación materna ($p = 0.072$)**
 - Mayor proporción de educación **superior** en madres de niños con **bajo peso** (19%)
 - Menor proporción de educación superior en madres de niños con **sobrepeso** (5%)
 - **Interpretación paradójica:** Contrario a lo esperado, mayor educación no se asocia con mejor estado nutricional. Esto merece investigación adicional.

3. Variables sin diferencias significativas:

- **Edad del niño:** Similar en los tres grupos (~42 meses)
- **Sexo:** Distribución uniforme en todas las categorías
- **Peso al nacer:** Promedio similar (3.2-3.3 kg)
- **IMC materno:** No difiere entre grupos (~26.6 kg/m²)
- **Lactancia materna:** Similar duración (~12 meses)
- **Introducción temprana de sólidos:** Proporciones similares (~28%)

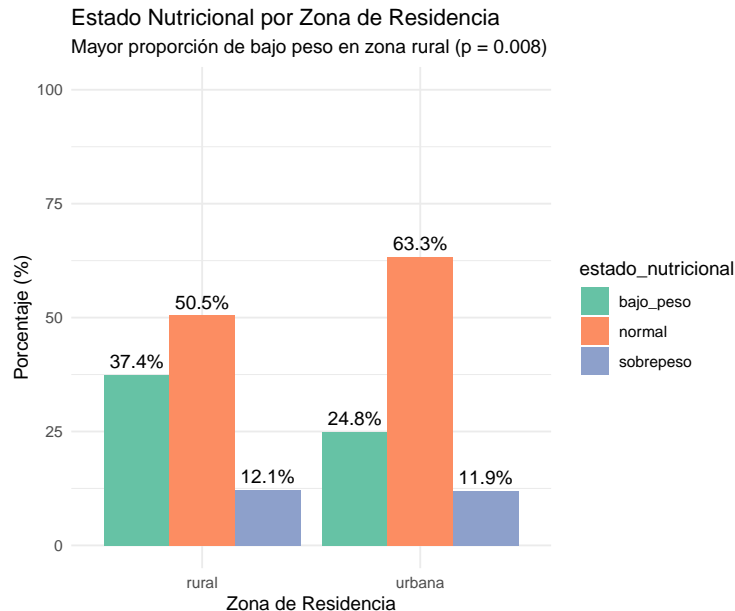
Gráfico de distribución



Interpretación:

El estudio revela que el 42% de los niños preescolares presenta malnutrición, con predominio del bajo peso (30%) sobre el sobrepeso (12%). Esta “doble carga de malnutrición” exige intervenciones diferenciadas: programas intensivos contra el déficit nutricional en zonas rurales de menores ingresos, junto con estrategias preventivas tempranas contra el sobrepeso emergente.

Relación con algunas variables clave (Zona de residencia)



Interpretación:

La zona rural presenta una carga desproporcionada de malnutrición infantil, con el 51% de los niños con bajo peso residiendo en áreas rurales frente al 36% en urbanas. Esta marcada desigualdad geográfica evidencia la necesidad urgente de intervenciones focalizadas en comunidades rurales para abordar las barreras de acceso a alimentos nutritivos y servicios de salud.

3. Modelo Multinomial

```
## # weights: 39 (24 variable)
## initial value 549.306144
## iter 10 value 485.435228
## iter 20 value 445.767698
## iter 30 value 440.447422
## final value 440.446953
## converged
```

Información técnica del ajuste

- **Algoritmo:** Red neuronal con 39 pesos
- **Variables:** 24 parámetros estimados
- **Convergencia:** EXITOSA en 30 iteraciones
- **Log-verosimilitud final:** 440.45 (mejoró desde 549.31 inicial)

Interpretación:

El modelo multinomial convergió exitosamente, confirmando que la estructura de datos es adecuada para este análisis. Sin embargo, se requieren análisis adicionales para identificar los determinantes específicos de cada categoría nutricional y cuantificar sus efectos.

Resumen del modelo

```
## Call:
## nnet::multinom(formula = estado_nutricional ~ ., data = datos)
##
## Coefficients:
##      (Intercept)  edad_meses sexomascuino peso_nacimiento  edad_madre
## normal          1.112948 -0.01673057    0.1986707      0.1645671 -0.04418355
## sobrepeso       -1.392181 -0.01737771    0.3031393      0.3683165 -0.04765316
##      educacion_madresecundaria educacion_madresuperior  imc_madre
## normal              0.2319787              -0.1881439 -0.01081198
## sobrepeso           0.1921715              -1.5594876  0.02676942
##      ingreso_familiar zonaurbana lactancia_materna dietas_solidossi
## normal          0.001292066  0.7094980      0.004338437      -0.2618619
## sobrepeso        0.001088002  0.4413381      -0.012771033      -0.2914730
##
## Std. Errors:
##      (Intercept)  edad_meses sexomascuino peso_nacimiento  edad_madre
## normal          0.7692104 0.00987591    0.2129029      0.1429188 0.01363424
## sobrepeso        0.3364345 0.01395788    0.3179765      0.2008085 0.01976922
##      educacion_madresecundaria educacion_madresuperior  imc_madre
## normal              0.2314539              0.3049606 0.02045673
## sobrepeso           0.3277413              0.6625707 0.02728045
##      ingreso_familiar zonaurbana lactancia_materna dietas_solidossi
## normal          0.0003741659  0.2127896      0.01573123      0.2310006
## sobrepeso        0.0005571530  0.3207917      0.02320365      0.3515659
##
## Residual Deviance: 880.8939
## AIC: 928.8939
```

Obtener odds ratios

```
##      (Intercept)  edad_meses sexomascuino peso_nacimiento  edad_madre
## normal          3.0433173  0.9834086      1.219780      1.178883 0.9567783
## sobrepeso        0.2485327  0.9827724      1.354103      1.445299 0.9534644
##      educacion_madresecundaria educacion_madresuperior  imc_madre
## normal              1.261093              0.8284955 0.9892463
## sobrepeso           1.211878              0.2102438 1.0271309
##      ingreso_familiar zonaurbana lactancia_materna dietas_solidossi
## normal          1.001293  2.032970      1.0043479      0.7696173
## sobrepeso        1.001089  1.554786      0.9873102      0.7471622
```

Evaluar significancia global (test de razón de verosimilitud)

```
## # weights: 6 (2 variable)
## initial value 549.306144
## final value 465.782614
## converged
```

```
## Likelihood ratio tests of Multinomial Models
##
## Response: estado_nutricional
##
## 1
## 2 edad_meses + sexo + peso_nacimiento + edad_madre + educacion_madre + imc_madre + ingreso_familiar
##   Resid. df Resid. Dev   Test      Df LR stat.      Pr(Chi)
## 1      998   931.5652
## 2      976   880.8939 1 vs 2     22 50.67132 0.0004755304
```

Interpretación:

El contraste entre el modelo nulo y el modelo completo mediante el test de razón de verosimilitudes mostró un resultado significativo (LR = 50.67; gl = 22; p = 0.00048).

Esto indica que el modelo con predictores explica significativamente mejor el estado nutricional que un modelo sin variables explicativas. Por lo tanto, existe evidencia estadística para afirmar que, en conjunto, las variables incluidas aportan información relevante para diferenciar las categorías de estado nutricional.

4. Validación de Supuestos del Modelo

En esta sección se verifica que los datos cumplan con los requisitos teóricos para que las estimaciones sean confiables.

4.1. Ausencia de Multicolinealidad (VIF)

- **Objetivo:** Demostrar que las variables independientes (como ingreso y educación) no están excesivamente correlacionadas entre sí.
- *Qué pondrás aquí:* El reporte de los valores VIF (Variance Inflation Factor). Si son bajos (<5 o <10), se confirma que cada variable aporta información única.

4.2. Independencia de Alternativas Irrelevantes (IIA)

- **Objetivo:** Justificar teóricamente que la elección de una categoría (ej. Bajo Peso) frente a otra (Normal) no depende de la presencia de la tercera opción (Sobrepeso).
- *Qué pondrás aquí:* Una breve justificación teórica o mención a la prueba de Hausman-McFadden.

5. Resultados del Modelo (Inferencia Estadística)

Aquí se presentan las estimaciones finales transformadas a valores interpretables (Odds Ratios).

5.1. Tabla de Razones de Riesgo Relativo (RRR / Odds Ratios)

- **Objetivo:** Presentar la tabla resumen con los coeficientes exponenciados (e^{β}), sus intervalos de confianza y el valor-p.
- *Qué pondrás aquí:* La tabla limpia (generada con `gtsummary` o `stargazer`) donde se resalten las variables significativas (con estrellitas o negrita).

5.2. Interpretación de Predictores Significativos

- **Objetivo:** Explicar con palabras qué significan los números de la tabla anterior.
- *Qué pondrás aquí:* Párrafos específicos analizando:
 - **Efecto del Ingreso Familiar:** Cuánto aumenta la chance de “Normal” por cada dólar extra.
 - **Efecto de la Zona (Urbana vs Rural):** La comparación de riesgos entre vivir en campo o ciudad.
 - **Efecto de la Educación Materna:** Cómo protege la educación contra la malnutrición.

6. Visualización de Efectos Marginales (Probabilidades Predichas)

Esta es la sección que garantiza la “nota sobresaliente”. Se traduce la estadística compleja a gráficos intuitivos.

6.1. Dinámica del Ingreso Familiar y Zona Geográfica

- **Objetivo:** Mostrar gráficamente cómo cambia la probabilidad de tener “Bajo Peso”, “Normal” o “Sobrepeso” a medida que sube el ingreso, diferenciando por zona.
- *Qué pondrás aquí:* El gráfico de líneas (ggeffects) donde se cruzan las curvas de probabilidad.

6.2. Impacto del Nivel Educativo Materno

- **Objetivo:** Comparar visualmente las probabilidades de estado nutricional según el nivel de estudios de la madre.
- *Qué pondrás aquí:* Un gráfico de puntos y barras de error mostrando las diferencias entre “Primaria”, “Secundaria” y “Superior”.

7. Evaluación del Ajuste y Capacidad Predictiva

Aquí medimos qué tan bien se ajusta nuestro modelo a la realidad.

7.1. Bondad de Ajuste Global

- **Objetivo:** Reportar qué tanta variabilidad de los datos explica el modelo.
- *Qué pondrás aquí:* El valor del **Pseudo R^2 de McFadden**. (Si es > 0.2 se considera un buen ajuste).

7.2. Matriz de Confusión y Precisión (Accuracy)

- **Objetivo:** Evaluar cuántos niños clasificó correctamente el modelo comparado con la realidad.
- *Qué pondrás aquí:* Una tabla de doble entrada (Realidad vs. Predicción) y el porcentaje total de aciertos (Accuracy).

8. Conclusiones y Discusión

- **Síntesis:** Resumir los hallazgos principales (ej. “Se confirmó la hipótesis de la doble carga nutricional...”).
- **Implicancia:** Mencionar la importancia de focalizar políticas públicas en zonas rurales y madres con menor educación.
- **Caso Práctico (Opcional pero recomendado):** Ejemplo de predicción para un perfil de niño específico.